11-51
371723

FINAL TECHNICAL REPORT-NASA GRANT-NAG5-4004
December 1, 1996-January 31, 1998

A: BACKGROUND

It is generally believed that an RNA World existed at an early stage in the history of life. During this early period, RNA molecules are seen to be potentially involved in both catalysis and the storage of genetic information. It is widely believed that this RNA World was extensive and therefore a sophisticated nucleic acid replication machinery would presumably predate the translation machinery which would not be needed until later stages in the development of life. This view of an extended RNA World is not necessarily correct. One might alternatively envision (Smith & Fox, 1995; Microbiologia SEM 11: 217-224, 1995), an abbreviated RNA World where peptide synthesis commenced very early and translation essentially co-evolved with replication. From the point of view of exobiology, the difference in these two views mainly affects the significance of studies of the extent of catalysis possible by RNA. In either case, the origin of the translation machinery and the principles of RNA evolution remain central problems in exobiology.

Translation presents several interrelated themes of inquiry for exobiology. First, it is essential, for understanding the very origin of life, how peptides and eventually proteins might have come to be made on the early Earth in a template directed manner. Second, it is necessary to understand how a machinery of similar complexity to that found in the ribosomes of modern organisms came to exist by the time of the last common ancestor (as detected by 16S rRNA sequence studies). Third, the RNAs that comprise the ribosome are themselves likely of very early origin and studies of their history may be very informative about the nature of the RNA World. Moreover, studies of these RNAs will contribute to a better understanding of the potential roles of RNA in early evolution.

The actual history of translation appears very complex. The problem is accentuated by the fact that a major portion of that history likely took place during a transition period before the final emergence of the last common ancestor as defined by the 16S rRNA phylogeny. Indeed, it is not unreasonable to suppose that it was the appearance of a sufficiently proficient translation machinery that was the final and decisive step in the emergence of true life. To date, the vast majority of work on translation has focused on function rather than the historical origins of the machinery.

It was the goal of his project to begin to address the history of translation with a focus on the RNAs. The approach was multifaceted. First, informatics was used to study what can be learned about the history of translation by the comparative studies of genomes. Second, direct experimental studies of rRNA evolution were carried out using 5S rRNA as a model system. The final component of the project was a direct, but highly speculative, assault on a key historical question: Can simple RNA molecules that might have arisen in an RNA world participate in peptide synthesis if they are carrying amino acids?

B: RESULTS

1. Bioinformatics

Two key issues were addressed in our bioinformatics study. In the first place we examined available complete bacterial genomes from the perspective of genome proximity. Genes that are next to each other are typically co-regulated, e.g. part of an operon. If the ribosomal machinery is indeed very ancient, one would expect that it would be among the first cellular systems to be subject to regulation. This was found to be overwhelmingly the case (Siefert et al., JME **45**:467-472, 1997). Comparisons of six genomes revealed that only 16 gene clusters, Table 1, involving 62 genes were conserved in all the genomes. Of these conserved gene clusters 12 contained ribosomal components representing 42 individual genes. Moreover, six of these ribosomal clusters are conserved in the Archaea as well. Subsequent publication of several additional genomes has allowed us to test this result further. Although some clusters ceased to be completely universal, the extensive conservation of these clusters to the exclusion of others is still noteworthy. These results tell us **(1) Regulation of the expression of genes associated with translation began before the Archaea/Bacteria split; and (2) Complex and nearly modern regulation of the transcription of the genes associated with transcriiption began very early**. There is in fact even more to the story however. A careful examination of the regulatory mechanisms associated with the earliest regulated gene clusters reveals in essentially every case where it is known (usually only in *Escherichia coli*) that the regulation involves RNA not DNA. The results thus strongly support an additional finding: **(3) An RNA genome, likely preceded the DNA Genome and that RNA genome encoded many if not all the conserved components of the modern translation machinery.**

The second bioinformatics project stems from the analysis of the 16S rRNA phylogeny. It is well known that by examining phenotypic properties from the perspective of this "tree of life" that one observes that all the earliest branchings are thermophilic, CO_2 utilizing, and anaerobic thus suggesting a phenotype for the last ancestor. We extended this "phylogeny mapping" approach to bacterial shape. The cell shape of several hundred Bacteria were examined and mapped onto the phylogeny with impressive results (Siefert and Fox, Microbiology, in press (1998)). **The primary finding is that there are persistent end state morphologies which seldom change once they are obtained**. Most notable of these was the coccus morphology which has independently, arisen numerous times in bacterial evolution with subsequent persistence. Other persistent morphologies include those of the spirochaetes and mycoplasmas. Overall, **the results strongly suggest that the common ancestor of the Bacterial line of descent was a rod and that the evolution of peptidoglycan was a fate sealing step in bacterial evolution.**

A third informatics effort has been to date less successful. It is reasonable to suppose that the much of the apparent complexity and size of the modern ribosome has its origins in large and small scale duplication events or fusion events. For example, it believed that initiation factor, IF-2, arose from EF-Tu in a duplication event that actually preceded the last common ancestor as defined by the 16S rRNA phylogeny. Likewise duplication and fusion events have been implicated by studies of ribosomal protein structure. The crystal structure of L14 reveals a clear internal duplication that again may predate the Archaea/(eu)Bacteria split. Crystal structures are known for at least nine ribosomal proteins. It has been found that S6, L7, L9 and L30 all contain a characteristic alpha-beta domain with an exposed beta sheet which is also exhibited by snRNP protein U1A. Likewise S17 and L14 both contain a five stranded beta-barrel structure. Given the similarities between some of the proteins examined to date it is not unreasonable to suppose that the entire set of the most universal ribosomal components actually arose from a significantly smaller number of ancestral proteins.

We therefore undertook an extensive comparison of the ribosomal proteins. Sequence alignments (http://www.bchs.uh.edu/~nzhou/evo_map/Evo_tran.htm) were generated for each of the most conserved ribosomal proteins. Initially a consensus sequence was constructed for each protein. These consensus sequences, in principle, are more like the ancestor and these were thus

Table 1. Conserved gene clusters[a]

Group	Short name, gene names	Type of regulation	Other organisms	Domain conservation
I	1. 16S, 23S, 5S rRNAs	RNA, DNA	*B. subtilis*, other Gram +, *Borrelia burgdorferi*	Bacteria/Archaea[g]
	2. S10[b] "operon" (*rps10, rpl3, rpl4, rpl23, rpl2, rps19, rpl22, rps3, rpl16, rpl29, rps17*)	RNA	*E. coli, B. subtilis, Thermotoga maritima*	Bacteria/Archaea[g]
	3. Str "operon" (*rps12, rps7, fusA*)	RNA	*E. coli, B. subtilis*	Bacteria/Archaea
	4. Spc[c] "operon" (*rpl14, rpl24, rpl5, rps14, rps8, rpl6, rpl18, rps5, rpl30, rpl15, secY*)	RNA	*E. coli, B. subtilis*	Bacteria/Archaea
	5. L13 "operon" (*rps9, rpl13*)	Not known	*E. coli, B. subtilis*	Bacteria/Archaea
	6. L11 "operon" (*rpl11, rpl1*)	RNA	*E. coli, B. subtilis, T. maritima*	Bacteria
	7. Alpha[d] "operon" (*rps13, rps11, rpoA, rpl17*)	RNA	*E. coli, B. subtilis*	Bacteria
	8. L35[e] "operon" (*infC, rpl35, rpl20*)	RNA	*E. coli, B. subtilis*	Bacteria
	9. L34 "operon" (*rpl34, rnpA*)	Not known	*E. coli, B. subtilis, B. burgdorferi*, other Gram+	Bacteria
	10. L21/L27 (*rpl21, rpl27*)	Not known	*E. coli*	Bacteria
	11. L10 "operon" (*rpl10, rpl12*)	RNA	*E. coli, B. subtilis, B. burgdorferi*, other Gram+	Bacteria/Archaea[g]
II	12. ATPases[f] (*atpB, atpE, atpF atpH, atpA, atpG, atpD, atpC*)	RNA	*B. subtilis*	Bacteria/Archaea
III	13. Beta "operon" (RNA polymerase) (*rpoC, rpoB*)	RNA	*B. subtilis*, other Gram+	Bacteria/Archaea
	14. Initiation factor (*nusA, infB*)	RNA	*B. subtilis, E. coli, Thermus aquaticus*	Bacteria
IV	15. Spermidine/putrescine Transport (*potA, potB, potC*)	Not known	*E. coli*	Bacteria
V	16. Chaperones (*groEL, groES*)	RNA, DNA	*B. subtilis*, other Gram+	Bacteria

[a] Conserved gene clusters identified in this study are numbered 1 through 16 and categorized in five groups according to gene function as studied in *E. coli*. Clusters 1–11 in group I are primarily RNA and protein constituents of the ribosome. Group II contains cluster 12, whose genes are involved in energy metabolism, e.g., the component of the ATP proton motive force interconversion enzyme, ATP synthase. The genes code for the hydrophilic F^1 unit which catalyzes the synthesis of ATP and membrane-bound hydrophobic F_0 unit, which forms the proton channel. Genes in group III are involved in RNA synthesis, modification, transcription, and translation. Cluster 13 codes for the β and β′ subunits of the DNA-dependent RNA polymerase. Cluster 14 codes for NusA, which modulates the rate of chain synthesis and IF-2, which binds tRNA to the ribosome complex during initiation of or protein synthesis. In addition it should be noted that cluster 7 includes the gene for the α subunit of DNA-dependent RNA polymerase. Cluster 15, in group IV, is a member of the superfamily of periplasmic binding-protein-dependent (BDP) and ATP-binding cassette (ABC) transporters, e.g., traffic ATPases which transport polyamines into the cell. Group V contains cluster 16, which codes for the molecular chaperones GroEL and GroES. The second column indicates whether the gene expression in the case of *E. coli* is regulated at the RNA or DNA level. The third column indicates other organisms in which the gene order is known to occur and the last column indicates the extent to which phylogenetic conservation of the gene order exists

[b] *rsp10* is not in this cluster in *Synechocystis* PCC 6803

[c] In *Synechocystis* PCC 6803 *rps14* has been relocated elsewhere and a homolog of *rpl30* has not been identified

[d] *rps4* is frequently found between *rps11* and *rpoA*

[e] *infC* is not located immediately upstream in *Synechocystis* PCC 6803

[f] The order in *Synechocystis* PCC 6803 is *atpH, atpG, atpF, atpD, atpA, atpC* with *atpB* and *atpE* located elsewhere

[g] The match with Archaea has exceptions or unusual features

intercompared to see if any of the proteins appeared to resemble one another, e.g. due to duplications. Clearly, millions of years of evolution may readily erase primary sequence similarity in nodal sequences which is why ancestral sequences were used. Despite this enhancement to the search, no convincing evidence of duplications was found. Duplicated domains may nevertheless continue to exhibit structural similarities even after primary sequence similarity is gone. In the absence of detailed structural information on all but a few of the ribosomal proteins, we compared predictions of structure. Although we were able to identify clusters of proteins that likely contain similar types of structure, e.g. primarily alpha or primarily beta, etc., we were unable to identify any examples of obviously similar folding. It seems likely that if the proposed duplication events did occur, that they will remain concealed until many additional high resolution ribosomal protein structures are determined.

2. RNA Evolution

The second component of the project continued efforts started under NASA grant NAGW-2108 to develop and utilize an experimental approach to study RNA evolution. The experimental system being used (Hedenstierna et al., Syst. Appl. Microbiol. **16**:280-286, 1993 and Lee et al., Origins Life & Evol. Biosphere **23**:365-372, 1993) examines the validity of variant 5S rRNA sequences in the vicinity of the modern *Vibrio proteolyticus* 5S rRNA sequence. This system has made it possible to conduct a detailed and extensive analysis of a local portion of the sequence space. This system allows us to explore a typical RNA sequence space, accessing both validity and invalidity of various sequence variants in the context of the *Escherichia coli* cellular environment Numerous mutants have been constructed during the last several years, and in excess of 135 *V. proteolyticus* derived constructs have been made and characterized. Data on many of these variants is available on our web page (http://www.bchs.uh.edu/~nzhou/temp/5snew.html). The vast majority of these constructs exhibit one of three major phenotypes. Type 1 constructs exhibit essentially wild type behavior and thus appear to be valid 5S rRNAs. Type 2 constructs do not accumulate to substantial levels in the cell. This apparently reflects the instability of the product as a result of either over processing originally or, less likely, premature degradation. These variants are found primarily in the region of the molecule that has been most strongly implicated in the binding of ribosomal protein L18. Type 3 constructs accumulate to very high levels but are absent from both 50S ribosomal subunits and 70S ribosomes. Type 3 mutations are especially common in the helix III/loop C subdomain of 5S rRNA which also is the region of greatest sequence conservation.

This RNA system is relevant to exobiology from several perspectives. First, is the reconstruction of ancestral sequences for key RNA molecules. Complete genome studies are now providing us with unprecedented amounts of information about extant sequences. This data can be used to construct hypothetical ancestral sequences. Traditionally, such ancestral sequences have been largely ignored as intermediates on the way to constructing phylogenetic trees. Are they actually meaningful predictions of the past? Studies by Steve Benner and others suggest that they are. Therefore, a promising new direction for studies on early life is to attempt to reconstruct likely ancestral components and study their properties. What are the rules for reconstructing ancestral sequences? Are there pitfalls we should know about? During the past funding period we began to examine these issues for RNA using our 5S rRNA system. The choice of 5S rRNA is a good choice in its own right as we will be able to address a second relevant objective, the evolutionary history of a highly conserved and early evolving component of the translation machinery. Finally, the results should be generalizable to understanding RNA sequence spaces in general and hence lead to a better understanding of a possible RNA world. During the course of this project we focused our efforts in three aspects of the problem:

(A) Can Particular Sequence Validity/Invalidity Be Predicted From Comparative Data?

If one fully understands a particular RNA shape space it should be possible to predict which sequences belong to it and which do not. Such a test would ideally be applied to possible ancestral sequences to insure their reasonableness. In attempting to make such predictions, three types of information would seem to be most relevant. The first requirement is structural information - preferably at atomic resolution for at least one sequence. The next item would be to have knowledge of how each point change individually effects validity in one sequence. Finally, comparative data-examples of valid sequences from as many naturally occurring organisms as possible in the local region of the shape space where predictions will be made. During the course of the project we sought principles that would facilitate predictions about various mutations basede on insights readily obtainable from comparative data..

One such principle is that mutations frequently accepted by closely related sequences are probably valid throughout much of an RNA sequence space. In the *Vibrio* cluster there are 15 positions that differ from the consensus sequence in at least 5 organisms and another 10 positions that differ in at least 3 positions. Six of these variant positions involve Watson-Crick base- pairs and in those cases the two changes always occur together. If one examines another cluster of sequences of similar phylogenetic diversity, e.g. a *Bacillus* cluster, one similarly sees variable positions but not the same ones. One's intuition is that these variable positions are "in play" and that these consistently seen changes will usually be valid throughout at least the local region of sequence space covered by the organisms that define the cluster. Thus, in the case of the *Vibrio* cluster one would intuitively surmise that when known "Vibrio variants" are separately introduced into V. *proteolyticus,* that they would have a very high probability of resulting in valid sequences. During the past year we tested this hypothesis by completing the construction of all of the relevant mutants to test this hypothesis. With the exception of one change, all of these mutants supported the hypothesis. **These results support the hypothesis that changes which frequently occur within a local region of sequence space have a very high probability of being accepted in any functional sequence in that region of sequence space.**

A second principle would appear to be as follows: Variants that are invalid in the standard sequence will usually only be found in other members of the shape space if there is also a compensating mutation. *V. proteolyticus* variants that are invalid in the *E. coli* cellular milieu do not occur naturally in other sequences in the *Vibrio* cluster. However, some of these "unexpected" changes do occur in more distant organisms- e.g. *Pseudomonas*. Clearly they are valid in that distant contexts. What has happened to make them valid? One possibility is a favorable change somewhere else in the cell. We believe however that it is far more likely that a compensating change has occurred elsewhere in the 5S rRNA. We believe it is possible, even though there are perhaps 25-40 changes total, to use the available information to identify what the compensating change is. In looking at several examples we have found that (1) many of the changes are obviously correlated because of base-pairing and hence can be treated as single events; (2) many of the positions belong to the standard set of variable positions (i.e. see item 2A above) and hence are probably inconsequential and can be ignored. If the "unexpected" change occurs in several organisms one can finally compare the residual set of interesting changes in these several organisms and look for a correlated position. We have in fact found four pairs of positions this way that we believe will be interdependent. During the course of the project we have begun the construction of many of the variants needed for this purpose. They were, however, not all completed and thus this question has not been definitively answered as of yet.

(B) Can Parsimony Analysis Be Used to Predict Credible Ancestral Sequences?

In order to better understand parsimony predictions we initially examined all intermediate sequences along alternative trajectories between two different pairs of valid points in the 5S rRNA sequence space (Lee *et al.*, 1997). In the case of the trajectory from the V. *proteolyticus* wild type sequence to the node corresponding to the V. *alginolyticus* sequence. There are 24 apparently equal paths, only five of which traverse exclusively valid sequences. If one assumes that paths where all intermediates can be fixed in a population have a significant advantage, then a properly constructed evolutionary tree based on parsimony should minimize total mutational events while utilizing only valid paths. The tree problem is more than finding the best trajectory between pairs of extant sequences as that is clearly not the course evolution took. We therefore used a parsimony approach to construct 100 phylogenetic trees from 32 previously determined *Vibrio* 5S rRNA sequences. Two of the ancestral nodal sequences, node 4 and node 25, were predicted to have existed in the past by all 100 trees. Prior experimental studies of the validity of individual point mutations carried by these two nodal sequences suggested that one of the sequences, node 25, would not be valid because the change C70U would likely result in RNA instability. These two nodal sequences were constructed by site directed mutagenesis. These sequences, and two construction intermediates were tested for validity as 5S rRNAs in the *Escherichia coli* cellular environment using previously developed procedures. **As predicted by the point mutation data, node 4 was found to be potentially valid whereas the node 25 RNA was unstable as predicted.** An analysis of wild type 5S rRNA sequences that carry the change C70U suggests that simultaneous change at one or more of three positions can compensate in some unknown way for the deleterious effect. **The results demonstrate that when used in conjunction with insights from studies of extant molecules, parsimony may provide very realistic predictions of possible historical sequences.**

(C) Development of Rapid Mutant Screening System

Characterization of an RNA sequence space by point mutation has been found to be effective but extremely tedious. An alternative approach to exploring a sequence space are combinatorial approaches (*in vitro* selection) that are being employed by many investigators. These methods are very rapid but have the disadvantage that one only learns what works. If a sequence alternative is not selected then one can not be sure whether it is bad or simply was not tried. During the project period we have been attempting to create an alternative methodology of "in vivo" exploration of an RNA sequence space. The idea is as follows; A host strain would be created in which genomic 5S rRNA genes would be damaged to the extent that the cell would just barely grow. A large set of potentially compensating 5S rRNA genes, the "test sequences" would be placed on plasmids, one per plasmid, and used to transform this sick cell line. Transformants would be selected and would be of two kinds depending on the validity/invalidity of the test sequence; those in which rapid growth is restored and those in which it is not. Sequencing would then be used to determine what the actual mutations are. During the project, major progress was made in developing the cell lines needed to construct the proposed *in vivo* experimental system. A gene replacement strategy was used to delete the 5S rRNA gene from a number of the seven ribosomal operons in *E. coli*. In particular, we have created strains lacking B, E, BE, BD, BDH and BDHE in the essentially wild type EMG-2 *E. coli* host strain. Because *rrn*B contains two 5S rRNAs, the BDHE minus strain lacks five of the eight 5S rRNA genes (Ammons, Rampersand & Fox, manuscript in preparation). The growth rate of these strains is seriously impaired on rich media where additional functional ribosomes are advantageous. Under these conditions, depending on temperature, the doubling time of the BDHE minus strain is 50-100% greater than the wild type EMG-2 strain. When cells are grown on plates it is easily possible to time observations such that a mixed plate of the wild type and BDHE minus strain contains both large and small colonies. We are currently conducting detailed growth studies and ribosome characterizations on these knockout strains and expect the results to be of relevance to our understanding of the role of 5S rRNA in protein synthesis in *E. coli.*

3. Peptide bond formation in the absence of ribosomes

It is increasingly apparent that a central issue in understanding the origins of translation is to explain how RNA mediated peptide bond formation got started. The critical breakthrough eventually provided by RNA mediated peptide synthesis was the development of template based specificity. Based on what is now known about the tRNA/synthetase interaction and our best guesses of the history of these molecules it seems very reasonable to suppose that one possible scenario is that translation had its origins in the ability of minihelices charged with amino acids to synthesize peptide bonds as has been proposed by Paul Schimmel and his colleagues. Certainly the case for the availability of such mini/micro helix RNAs on the early Earth has recently been placed on much firmer ground. David Usher announced at a symposium in Saratoga Springs that short RNAs with appropriate linkages (i.e. potential minihelices) were formed in his "day-night" machine and last year Jim Ferris recently reported the synthesis of RNA oligomers up to 50 nucleotides on clay surfaces. If minihelix mediated peptide synthesis did in fact occur it was not necessarily template mediated and may thus have initially competed with alternative methods, e.g. synthesis on clay surfaces. When they first appeared, RNA/nucleotide linked amino acids or peptides may have simply extended the range of catalysis of ribozymes (coenzyme A may be a relic of such a period) or perhaps they were more compatible with membrane enclosure. Their ability to participate in peptide bond formation may have come to the fore later. The central goal of our work was to determine if minihelices can participate in peptide bond formation in the absence of ribosomes. Although we were not able to accomplish this with RNAs as small as minihelices during this project, considerable progress was made with somewhat larger RNAs.

The key to our approach was a very preliminary report by Shimizu (J. Biochem., 119:832-834, 1996) that the dipeptide alanylhistidine catalyzes dipeptide formation between two aminoacylated tRNAs in the absence of ribosomes. This paper had not been well received and is apparently generally regarded as suspect due to ill defined conditions on many of the experiments that were reported. In particular, it is not clear that the synthetases are always removed following charging. It therefore was possible that the peptide bond formation is in some instances actually being catalyzed by the synthetase. The attractiveness of a histidine dipeptide as a prebiotic catalyst and the existing evidence for an active role for histidine in the modern translation machinery prompted us to fully evaluate this report.

Our work has been conducted primarily with a leucine system. In order to charge RNAs we first needed to prepare purified leucine tRNA synthetase as these enzymes are not commercially available in pure form. This enzyme was produced by over expression from a plasmid, pLeuS-1, carrying the *E. coli* leucine tRNA synthetase gene and purified. In addition to the primary work described here, the availability of significant quantities of the enzyme allowed us to examine issues relating to the fidelity of amino acid recognition by synthetases (Martinis et al., *Nucl. Acids Res. Symp. Ser.* **36:** 125-128, 1997).

In our first series of peptide synthesis experiments, panel 1-3 of Figure 1, we examined incubations containing either charged leucine tRNA, a mixture of crude tRNAs (i.e. including small amounts of leucine tRNA) or a control sample without tRNA, with and without ala-his. The reactions were incubated at 37°C for one hour with [^{14}C]-leu-tRNA^{leu}. Radiolabeled leu-leu products and leu were hydrolyzed from tRNA^{leu} by treatment with 0.2N KOH, separated on TLC plates, and then visualized using a Fuji BAS 1000 phosphoroimager. In the first lane of each TLC plate is a non-hydrolyzed control. The [^{14}C]-leucine in this case remains covalently bound to the untreated tRNA and thus stays essentially at the origin. This figure shows that in the presence of ala-his that a second product with the mobility of leu-leu is formed. The identity of this product was verified by mass spectroscopy. **We have demonstrated unequivocally that peptide bond formation can be achieved with an RNA/peptide complex in the absence of ribosomes and tRNA synthetase.** The addition of a short template containing several leucine codons does not improve the reaction. When the template is included and the ala-his dipeptide is omitted no leucine dipeptide is formed. Thus, unlike Shimizu **we find the reaction to be template independent** (he used phe, lys, pro and gly tRNAs). In the second panel only small amounts of charged leucine tRNA are present and the dipeptide spots are extremely faint. In the complete absence of charged leucine tRNA, panel 3, as expected nothing is seen. This experiment has been successfully repeated several times.

The purpose of the second series of experiments, Figure 2, was twofold; (1) to confirm the claim that tRNA synthetase alone could catalyze peptide bond formation and (2) assuming it could- to establish that the purification procedures we were using to separate charged tRNA from its synthetase were sufficient to prevent peptide bond formation by the synthetase alone. In these experiments the synthetase alone (no tRNA) is incubated with leucine, Mg^{++}, and ATP under conditions in which the leucine adenylate can form. In panel A, the reaction time was 30 minutes and in panel B it was 5 hours. In each case, the sample was either run directly or first "purified" by alcohol precipitation and phenol extraction. In lanes marked ala-his positive, the dipeptide was added after the "purification" step if any. Under short reaction times no dipeptide product is seen whereas under longer reaction times considerable amounts do indeed accumulate. When the dipeptide is formed it appears with or without the ala-his dipeptide which if anything is inhibiting the reaction. In all cases the "purification" procedure clearly prevented the reaction, presumably by eliminating both synthetase and free leucine. **The results confirm that tRNA synthetase alone catalyzes synthesis of leucine dipeptide with high yield.** Depending on the history of this enzyme which is not yet fully known this may or may not have significance to early life.

In the third series of experiments shown in Figure 3, we examined the effect of pH on the dipeptide reaction It is seen that the best yield of leucine dipeptide, 13%, occurred at pH=7, What is especially interesting in this case is the appearance of a new spot at low pH's that was never seen by Shimuizu. This product was shown to be an alternate form of Leu-Leu dipeptide by mass spectroscopy rather than a reaction intermediate. We have also examined the effect of alternative peptide catalysts on the reaction (data not shown) to test the specificity of ala-his. No peptide bond formation occurred with his-ala, ala-ala; val-asp, or phe-phe. In addition, we examined the effect of histidine alone and imidazole alone on the reaction. In each case no dipeptide was formed. Apparently ala-his is a better leaving group than either histidine or imidazole at pH 7.

ALA-HIS as a Catalyst for Non-Ribosomal Peptide Bond Formation

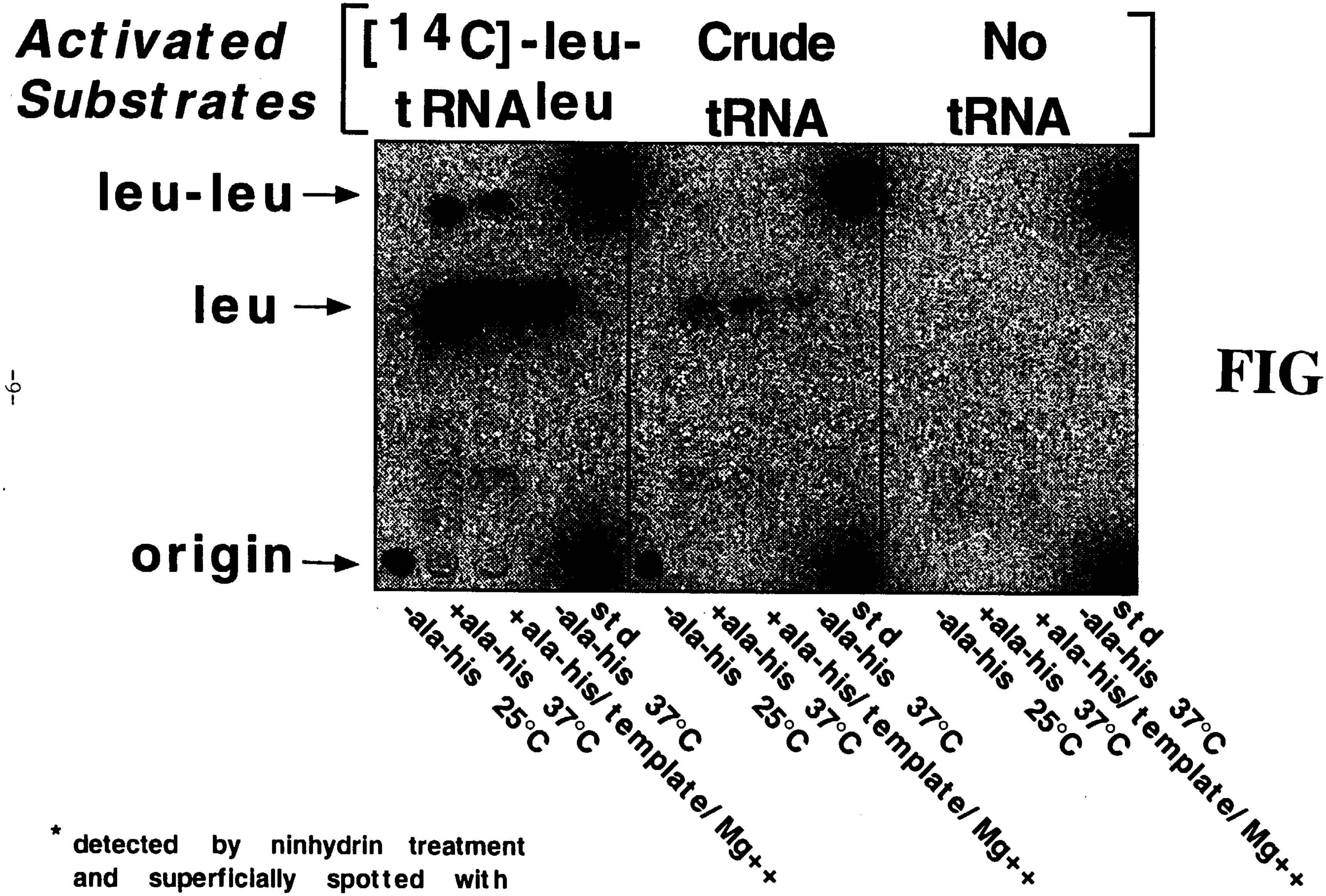

FIGURE 1

* detected by ninhydrin treatment and superficially spotted with a [^{14}C] radiolabel.

ALA-HIS Catalyzed Non-Ribosomal Peptide Bond Formation (pH Variations)

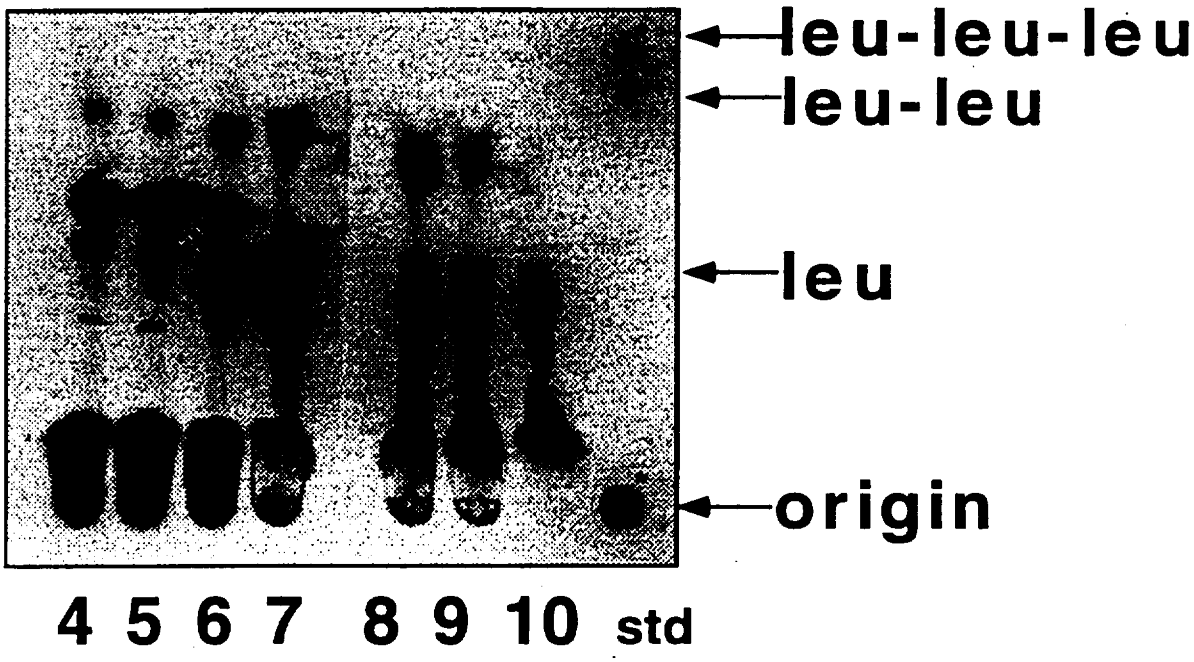

Figure 3: pH effect on leu-leu non-ribosomal dipeptide formation in the presence of the dipeptide catalyst Ala-His.

Ala-His interactions with [^{14}C]-leu-tRNA^{Leu} in a range of acidic to basic enviroments produce possible intermediates or by-products. At lower pH values, several unknown compounds are formed and hydrolysis of the tRNA by 0.2N KOH treatment is less efficient. At pH 7, a 13% yield of leu-leu dipeptide is obtained.

In addition, preparations for an attempt to obtain peptide bond formation with minihelices have also been made. In order to facilitate synthesis of test RNAs we purified T7 polymerase and used it to generate leucine minihelix RNA by runoff transcription from a DNA template. To date we have been unable to charge minihelices for leucine and we are now preparing the materials needed to attempt these experiments with an alternative alanine system for which minihelices are known to charge. We were successful in charging a leu-tRNA in which the anticodon had been deleted. Thus, **we have demonstrated that the peptide synthesis reaction does not require an anticodon in the RNA.**

These results demonstrate that RNA/peptide complexes resembling what are found in ribosomes today could have been involved in peptide bond formation in prebiotic era before the entire ribosome had evolved. For the first time we have a reasonable starting point for the evolution of the protein synthesis machinery. The catalyst required, a di-peptide is clearly compatible with what was available in the prebiotic world but the RNA component is rather large, hence the interest in showing that a charged minihelix would work. Assuming that a minihelix of less that 30 residues works, we will have demonstrated a very feasible peptide bond system for the prebiotic world. Several issues will remain however. First, how would small RNAs come to be aminoacylated in the prebiotic world? It has been shown that RNAs can catalyze this reaction but a more realistic solution would be preferable. Second, what would have prevented hydrolysis of the di-peptide product from the RNA? Once hydrolysis occurs one can not continue on to make larger peptides. Finally, how/why would the reaction have become template directed? Although much has been accomplished in this line of investigation, further work will clearly be required.

C: DISSEMINATION OF RESULTS

Dissemination of results to the larger scientific community is an important activity of any scientific research project. To this end we have utilized the traditional approaches of publication, meeting posters & presentations and invited lectures.
In addition, we have begun construction of a world wide web site where some of our results are also presented (http://www.bchs.uh.edu/~nzhou/temp/5snew.html).
Our efforts to the knowledge obtained were as follows:

1. Publications

Pitulle, C., DSouza,L., and Fox, G. E. "A Low Molecular Weight Artificial RNA of Unique Size with Multiple Probe Target Regions", *Syst. Appl. Microbiol.* **20:** 133-136 (1997).

Lee, Y-H., DSouza, L. M. and Fox, G. E. "Equally Parsimonious Pathways Through an RNA Sequence Space are not Equally Likely" , *J. Mol. Evol.,* **45:** 278-284 (1997).

Siefert, J. L., Martin, K. A., Abdi, F., Widger, W. R., and Fox, G. E. "Conserved Gene Clusters in Bacterial Genomes Provide Further Support for the Primacy of RNA", *J. Mol. Evol.,* **45:** 467-472 (1997).

Martinis, S. A. and Fox, G. E. "Non-standard Amino Acid Recognition by *Escherichia coli* Leucyl-tRNA Synthetase", *Nucl. Acids Res. Symp. Ser.* **36:** 125-128 (1997).

Siefert, J. L. and Fox, G. E., "Phylogenetic Mapping of Bacterial Morphology", *Microbiology,* in press, (1998).

Ammons, D., Rampersad, J., and Fox, G. E. "A Genomically Modified Marker Strain of *Escherichia coli, Curr. Microbiol.*, in press, (1998).

2. Invited Presentations

"Exploration of RNA Sequence Spaces" Invited Seminar, Department of Ecology and Evolutionary Biology; Rice University, Houston, Tx., December 9, 1996.

Invited Lecture "Progenotes and Archaebacteria: Facts and Hypotheses about Earliest Life"; in public lecture series entitled Origin of Life: Earth, Mars and Beyond; February 25, 1997; Houston, Texas.

Participant on Discussion Panel at end of Conference on Early Mars- Sponsored by Lunar & Planetary Institute, Houston, Texas, April 24-27, 1997.

Invited Sigma Chi Lecture, "Tracking *E. coli*: Artificial Stable RNAs in Bacterial Monitoring", Centers for Disease Control, Atlanta, Ga., October 30, 1997.

Invited Symposium Speaker, Texas American Society of Microbiology Branch Meeting, Houston, Texas, November 6-7, 1997.

Invited Speaker, "The Origins of the Translation Apparatus", 6th Symposium on Chemical Evolution and the Origin and Evolution of Life, NASA-Ames Research Center, November 17-20, 1997.

Invited Speaker (Susan Martinis), "Primordial Mechanisms of Biological Protein Synthesis", University of Texas Health Science Center, Houston, Tx., February 12, 1998.

3. Abstracts and Meeting Presentations

Martinis, S. A. and Fox, G. E.; "Amino Acid Recognition by Leucyl-tRNA Synthetase", 17th International tRNA Workshop, Kisarazu City, Japan, May 10-15, 1997.

Dsouza, L. M., Lee, Y-H., Pitulle, C., and Fox, G. E.; "5S rRNA as a Tool for Studying Molecular Evolution" Abstracts- RNA '97, p212, Banff, Alberta, Canada , May 27-June 1, 1997,

Larkin, D., Martinis, S.A., and Fox, G. E.; "Primordial Protein Synthesis Systems" NATO/EMBO/FEBS Advanced Study Institute: Biomolecular Recognition, Spetsai, Greece, September 1-14, 1997.

Martinis, S. A. and Fox, G. E.; "Non-standard Amino Acid Recognition by E. coli Leucyl-tRNA Synthetase", Symposium on RNA Biology II - RNA: Tool and Target, Research Triangle Park, NC, October 17-19, 1997.

Larkin, D. C., Martinis, S. A., and Fox, G. E., "The Origins of the Translation Machinery", Abstracts Sixth Symposium on Chemical Evolution and the Origin and Evolution of Life, p54, NASA Ames Research Center, Mountain View, CA., November 17-20, 1997.
Martinis, S. A. and Fox, G. E., "Fidelity of Amino Acid Recognition by tRNA Synthetases" Abstracts Third Annual Meeting RNA Society, p. 457, Madison, WI, May 26-31, 1998.

Zhang, Z., Dsouza, L.M., Lee, Y-H., Yang, Y., and Fox, G. E., "Variable Positions May Be Evolutionarily Unconstrained Over a Large Range of Sequence Contexts", Abstracts Third Annual Meeting RNA Society, p. 750, Madison, WI, May 26-31, 1998.